

# Wald type and Phi-divergence based test-statistics for isotonic binomial proportions

Martin, N.<sup>1</sup>, Mata, R.<sup>2</sup> and Pardo, L.<sup>2</sup>

<sup>1</sup>Dep. Statistics, Carlos III University of Madrid, 28903 Getafe (Madrid), Spain

<sup>2</sup>Dep. Statistics and O.R., Complutense University of Madrid, 28040 Madrid, Spain

February 28, 2014

## Abstract

In this paper new test statistics are introduced and studied for the important problem of testing hypothesis that involves inequality constraint on proportions when the sample comes from independent binomial random variables: Wald type and phi-divergence based test-statistics. As a particular case of phi-divergence based test-statistics, the classical likelihood ratio test is considered. An illustrative example is given and the performance of all of them for small and moderate sample sizes is analyzed in an extensive simulation study.

*Keywords and phrases:* Wald-type statistics, Phi-divergence statistics, Inequality constraints, Loglinear model, Logistic regression.

## 1 Introduction

Ordinal categorical data appear frequently in the biomedical research literature, for example, in the analysis of  $I$  independent binary random variables related to an increasing ordered categorical variable. It is well-known that for such data it is not possible to use the classical test-statistics such as chi-square or likelihood ratio with chi-squared asymptotic distribution, but there exist appropriate order-restricted test-statistics with chi-squared-bar asymptotic distribution. To illustrate this problem a modification of an example given in Silvapulle and Sen (2005) is considered in this introductory section.

Table 1 contains a subset of data from a prospective study of maternal drinking and congenital malformations. Women completed a questionnaire, early in their pregnancy, concerning alcohol use in the first trimester; complete data and details are available elsewhere (Graubard

| $i$ | $n_i$ | $n_{i1}$ | $n_{i2}$ |
|-----|-------|----------|----------|
| 1   | 17114 | 48       | 17066    |
| 2   | 14502 | 38       | 14464    |
| 3   | 793   | 5        | 788      |
| 4   | 165   | 2        | 163      |

Table 1: Number of individuals with ( $j=1$ ) and without ( $j=0$ ) congenital sex-organ malformation cross-classified according to the maternal alcohol consumption level ( $i=1,2,3,4$ ).

and Korn, 1987). Specifically, women were asked what was the amount of alcohol taken during the first three months of their pregnancy and four categories of drink doses are considered ( $I = 4$ ), no alcohol consumption ( $i = 1$ ), average number of alcoholic drinks per day less than one but greater than zero ( $i = 2$ ), one or more and less than three alcoholic drinks per day ( $i = 3$ ) and three or more alcoholic drinks per day ( $i = 4$ ). In terms of a binary random variable with  $n_i$  individuals in total (see the second column in Table 1) with independent behavior with respect to having congenital malformations, the individuals not having congenital malformations are considered to be unsuccessful ( $j = 2$ , see the last column in Table 1) and successful otherwise ( $j = 1$ , see the third column in Table 1). Let  $\pi_i$  be the probability of a success associated with the  $i$ -th alcohol dose. Let us consider some statistical inference questions that may arise in this example and in similar ones with binomial probabilities.

1. Is there any evidence of maternal alcohol consumption being related to malformation of sex organ? To answer this question, the null and alternative hypotheses may be formulated as

$$H_0 : \pi_1 = \pi_2 = \pi_3 = \pi_4 \text{ vs. } H_1 : \pi_1, \pi_2, \pi_3, \pi_4 \text{ are not all equal,}$$

respectively. However, this formulation is unlikely to be appropriate because the main issue of interest is the possible increase in the probability of malformation as alcohol consumption increases.

2. Is there any evidence that an increase in maternal alcohol consumption is associated with an increase in the probability of malformation?. This question, as it stands, is quite broad to give a precise formulation of the null and the alternative hypotheses. One possibility is to formulate the problem in the following way,

$$H_0 : \pi_1 = \pi_2 = \pi_3 = \pi_4 \text{ vs. } H_1 : \pi_1 \leq \pi_2 \leq \pi_3 \leq \pi_4 \text{ with at least one inequality being strict.} \quad (1)$$

Consider an experiment with  $I$  increasing ordinal categories for a variable  $X$ . Suppose that  $n_i$  prefixed individuals are assigned to the  $i$ -th category and  $n = \sum_{i=1}^I n_i$ . The individuals are

followed over time for the development of an event of interest  $Y$  and the events related to the individuals are independent. Let  $N_{i1}$  be the random variable that represents the number of individuals related to successful events ( $Y = 1$ ) out of the total assigned to the  $i$ -th category,  $n_i$ ,  $i = 1, \dots, I$ . If we denote by  $\pi_i = \Pr(Y = 1|X = i)$  the probability of a success associated with the  $i$ -th category, we have that  $N_{i1}$  is a Binomial random variable with parameters  $n_i$  and  $\pi_i$ ,  $i = 1, \dots, I$ . Let  $N_{i2}$  denote the number of unsuccessful events associated with the  $i$ -th category, i.e.  $N_{i2} = n_i - N_{i1}$ , then the contingency table of a realization of  $(N_{i1}, N_{i2})$ ,  $i = 1, \dots, I$ , is in the last two columns of the following table

|          |          |                         |
|----------|----------|-------------------------|
| $n_1$    | $n_{11}$ | $n_{12} = n_1 - n_{11}$ |
| $\vdots$ | $\vdots$ | $\vdots$                |
| $n_i$    | $n_{i1}$ | $n_{i2} = n_i - n_{i1}$ |
| $\vdots$ | $\vdots$ | $\vdots$                |
| $n_I$    | $n_{I1}$ | $n_{I2} = n_I - n_{I1}$ |

.

Our purpose in this paper is to propose new order-restricted test statistics, Wald-type and phi-divergence based test-statistics for testing

$$H_0 : \pi_1 = \pi_2 = \dots = \pi_I, \tag{2}$$

$$H_1 : \pi_1 \leq \pi_2 \leq \dots \leq \pi_I \text{ with at least one inequality being strict.}$$

The classical likelihood ratio test statistic will appear as a particular case of phi-divergence based test-statistics. A log-linear formulation of (1) is proposed in Section 2, fundamental for defining the Wald type test-statistics. In Section 3 the families of phi-divergence test statistics are presented. Section 4 is devoted to solve the problem presented in this Section 1 for a illustrative example. An extensive simulation study is carried out in Section 5.

## 2 Formulation for isotonic binomial proportions in terms of log-linear and logistic regression modeling: Wald type test-statistics

Reparametrizing the initial problem through log-linear modeling, the formulation of the null hypothesis is strongly simplified since all the interaction parameters are zero under the null

hypothesis and this is appealing, in special, to create Wald type test-statistics. Let

$$\begin{aligned}\mathbf{p} = \mathbf{p}(\boldsymbol{\theta}) &= (p_{11}(\boldsymbol{\theta}), p_{12}(\boldsymbol{\theta}), p_{21}(\boldsymbol{\theta}), p_{22}(\boldsymbol{\theta}), \dots, p_{I1}(\boldsymbol{\theta}), p_{I2}(\boldsymbol{\theta}))^T \\ &= \left(\frac{n_1}{n}\pi_1, \frac{n_1}{n}(1 - \pi_1), \frac{n_2}{n}\pi_2, \frac{n_2}{n}(1 - \pi_2), \dots, \frac{n_I}{n}\pi_I, \frac{n_I}{n}(1 - \pi_I)\right)^T\end{aligned}\quad (3)$$

be the probability vector of the following saturated log-linear model

$$\log p_{ij}(\boldsymbol{\theta}) = u + u_{1(i)} + \theta_{2(j)} + \theta_{12(ij)}, \quad (4)$$

with

$$u_{1(I)} = 0, \quad \theta_{2(2)} = 0, \quad \theta_{12(i2)} = 0, i = 1, \dots, I - 1, \quad \theta_{12(Ij)} = 0, j = 1, 2, \quad (5)$$

being the identifiability constraints,

$$\boldsymbol{\theta} = (\theta_{2(1)}, \theta_{12(11)}, \dots, \theta_{12(1I-1)})^T \quad (6)$$

the unknown parameters vector and  $\mathbf{u} = (u, u_{1(1)}, \dots, u_{1(I-1)})^T$  with

$$u = u(\boldsymbol{\theta}) = \log \frac{n_I/n}{1 + \exp\{\theta_{2(1)}\}}, \quad (7)$$

$$u_{1(i)} = u_{1(i)}(\boldsymbol{\theta}) = \log \frac{\frac{n_i}{n_I} (1 + \exp\{\theta_{2(1)}\})}{1 + \exp\{\theta_{2(1)} + \theta_{12(i1)}\}}, \quad i = 1, \dots, I - 1, \quad (8)$$

the redundant parameters, obtained through  $\boldsymbol{\theta}$  taking into account  $p_{i1}(\boldsymbol{\theta}) + p_{i0}(\boldsymbol{\theta}) = \frac{n_i}{n}$ ,  $i = 1, \dots, I$ . In terms of the log-linear formulation, (2) is equivalent to

$$H_0 : \theta_{12(11)} = \theta_{12(21)} = \dots = \theta_{12(I-1,1)} = 0, \quad (9)$$

$$H_1 : \theta_{12(11)} \leq \theta_{12(21)} \leq \dots \leq \theta_{12(I-1,1)} \leq 0 \text{ with at least one inequality being strict.}$$

Notice that  $\theta_{2(1)}$  is a nuisance parameter since it does not interfere in (9). In particular, under the null hypothesis of  $\pi_0 = \pi_1 = \pi_2 = \dots = \pi_I$ , the value of the nuisance parameter is  $\theta_{2(1)} = \text{logit}(\pi_0) = \log[\pi_0/(1 - \pi_0)]$ , and thus it contains all the information about the homogeneous probability vector.

In matrix notation, we can express the vector of parameters of the log-linear model in terms of the following logistic regression

$$\text{logit}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\theta} \quad (10)$$

where

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_I)^T, \quad \mathbf{X} = \begin{pmatrix} \mathbf{1}_{I-1} & \mathbf{I}_{I-1} \\ 1 & \mathbf{0}_{I-1}^T \end{pmatrix},$$

$\mathbf{I}_a$  is the identity matrix of order  $a$ ,  $\mathbf{1}_a$  is the  $a$ -vector of ones and  $\mathbf{0}_a$  is the  $a$ -vector of zeros. Since a saturated model has been considered,  $\mathbf{X}$  is a full rank matrix and thus we can consider

$$\boldsymbol{\theta} = \mathbf{X}^{-1} \text{logit}(\boldsymbol{\pi}), \quad (11)$$

and on the other hand (9) in matrix notation is given by

$$H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{0}_{I-1}, \quad (12)$$

$$H_1 : \mathbf{R}\boldsymbol{\theta} \leq \mathbf{0}_{I-1} \text{ and } \mathbf{R}\boldsymbol{\theta} \neq \mathbf{0}_{I-1},$$

with  $\mathbf{R} = (\mathbf{0}_{I-1}, \mathbf{G}_{I-1})$ , and  $\mathbf{G}_{I-1}$  is a square matrix of order  $I-1$  with 1-s in the main diagonal and  $-1$ -s in the upper superdiagonal.

We shall consider three parameter spaces for  $\boldsymbol{\theta}$

$$\Theta_0 = \{\boldsymbol{\theta} \in \mathbb{R}^I : \mathbf{R}\boldsymbol{\theta} = \mathbf{0}_{I-1}\} \subset \tilde{\Theta} = \{\boldsymbol{\theta} \in \mathbb{R}^I : \mathbf{R}\boldsymbol{\theta} \leq \mathbf{0}_{I-1}\} \subset \Theta = \mathbb{R}^I,$$

i.e. while  $\Theta$  is the restricted parameter space,  $\tilde{\Theta}$  is unrestricted, and  $\Theta_0$  becomes the parameter space under the null hypothesis. It is well known that, the Fisher information matrix for  $\boldsymbol{\theta} \in \Theta$  in the logistic regression is given by

$$\mathcal{I}_F(\boldsymbol{\theta}) = \mathbf{X}^T \text{diag}\{\nu_i \pi_i (1 - \pi_i)\}_{i=1}^I \mathbf{X}, \quad (13)$$

where  $\nu_i = \lim_{n \rightarrow \infty} \frac{n_i}{n}$ ,  $i = 1, \dots, I$ . The following result provides the explicit expression of the Fisher information matrix under the null hypothesis given in (2) or (9).

**Theorem 1** For  $\boldsymbol{\theta}_0 \in \Theta_0$  in the model (4) or (10), the Fisher information matrix is given by

$$\mathcal{I}_F(\hat{\boldsymbol{\theta}}) = \pi_0(1 - \pi_0) \begin{pmatrix} 1 & \nu_1 & \nu_2 & \cdots & \nu_{I-1} \\ \nu_1 & \nu_1 & 0 & \cdots & 0 \\ \nu_2 & 0 & \nu_2 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \nu_{I-1} & 0 & 0 & \cdots & \nu_{I-1} \end{pmatrix}. \quad (14)$$

**Proof.** It is immediate by plugging  $\pi_i = \pi_0$ ,  $i = 1, \dots, I$ , to (13). ■

If  $\hat{\boldsymbol{\theta}}$ ,  $\tilde{\boldsymbol{\theta}}$  and  $\bar{\boldsymbol{\theta}}$  represent the maximum likelihood estimator (MLE) of  $\boldsymbol{\theta}$  focussed on the

parameter spaces,  $\Theta_0$ ,  $\tilde{\Theta}$ ,  $\Theta$  respectively, according to Silvapulle and Sen (2005, pages 154 and 166) we can consider three Wald-type test-statistics,

$$W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) = n\tilde{\boldsymbol{\theta}}^T \mathbf{R}^T \left( \mathbf{R} \mathcal{I}_F^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{R}^T \right)^{-1} \mathbf{R} \tilde{\boldsymbol{\theta}}, \quad (15)$$

$$H(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) = n(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^T \mathcal{I}_F(\hat{\boldsymbol{\theta}}) (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}), \quad (16)$$

$$D(\bar{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) = n(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^T \mathcal{I}_F(\hat{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) - n(\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^T \mathcal{I}_F(\hat{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}), \quad (17)$$

where

$$\begin{aligned} \mathcal{I}_F(\hat{\boldsymbol{\theta}}) &= \hat{\pi}_0(1 - \hat{\pi}_0) \begin{pmatrix} 1 & (\hat{\boldsymbol{\nu}}^*)^T \\ \hat{\boldsymbol{\nu}}^* & \text{diag}(\hat{\boldsymbol{\nu}}^*) \end{pmatrix}, \\ \mathcal{I}_F(\tilde{\boldsymbol{\theta}}) &= \mathbf{X}^T \text{diag}\{\hat{\nu}_i \tilde{\pi}_i(1 - \tilde{\pi}_i)\}_{i=1}^I \mathbf{X}, \\ \hat{\boldsymbol{\nu}}^* &= (\hat{\nu}_1, \dots, \hat{\nu}_{I-1})^T = \left( \frac{n_1}{n}, \dots, \frac{n_{I-1}}{n} \right)^T. \end{aligned}$$

These test-statistics, have according to Proposition 4.4.1 in Silvapulle and Sen (2005), the same asymptotic distribution as the likelihood ratio test-statistic.

**Proposition 2** *Under the null hypothesis given in (2) or (9), the expression of  $W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$ , given in (15), is as follows*

$$W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) = n\hat{\pi}_0(1 - \hat{\pi}_0)(\tilde{\boldsymbol{\theta}}^*)^T \boldsymbol{\Sigma}_{\hat{\boldsymbol{\nu}}^*} \tilde{\boldsymbol{\theta}}^*,$$

where  $\boldsymbol{\theta}^* = (\theta_{12(11)}, \dots, \theta_{12(1I-1)})^T$ ,  $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\nu}}^*} = \text{diag}(\hat{\boldsymbol{\nu}}^*) - \hat{\boldsymbol{\nu}}^*(\hat{\boldsymbol{\nu}}^*)^T$ ,

$$\begin{aligned} \hat{\pi}_0 &= \frac{N_{\bullet 1}}{n}, \\ N_{\bullet 1} &= N_{11} + N_{21} + \dots + N_{I1}. \end{aligned}$$

**Proof.** From Theorem 1 the  $2 \times 2$  block structure of the Fisher information matrix is

$$\mathcal{I}_F(\boldsymbol{\theta}_0) = \hat{\pi}_0(1 - \hat{\pi}_0) \begin{pmatrix} 1 & (\hat{\boldsymbol{\nu}}^*)^T \\ \hat{\boldsymbol{\nu}}^* & \text{diag}(\hat{\boldsymbol{\nu}}^*) \end{pmatrix}$$

and

$$\mathcal{I}_F^{-1}(\boldsymbol{\theta}_0) = \frac{1}{\hat{\pi}_0(1 - \hat{\pi}_0)} \begin{pmatrix} a & \mathbf{b}^T \\ \mathbf{c} & \mathbf{D} \end{pmatrix},$$

with  $\mathbf{D} = \boldsymbol{\Sigma}_{\hat{\boldsymbol{\nu}}^*}^{-1}$ ,  $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\nu}}^*} = \text{diag}(\hat{\boldsymbol{\nu}}^*) - \hat{\boldsymbol{\nu}}^*(\hat{\boldsymbol{\nu}}^*)^T$ . But

$$\mathbf{R} \mathcal{I}_F^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{R}^T = \frac{1}{\hat{\pi}_0(1 - \hat{\pi}_0)} \mathbf{G}_{I-1} \boldsymbol{\Sigma}_{\hat{\boldsymbol{\nu}}^*}^{-1} \mathbf{G}_{I-1}^T$$

and

$$R\tilde{\theta} = G_{I-1}\theta^*.$$

Therefore,

$$\begin{aligned} W(\tilde{\theta}, \hat{\theta}) &= n\tilde{\theta}^T R^T \left( R\mathcal{I}_F^{-1}(\hat{\theta})R^T \right)^{-1} R\tilde{\theta} \\ &= n\hat{\pi}_0(1 - \hat{\pi}_0)(\tilde{\theta}^*)^T \Sigma_{\hat{\nu}^*} \tilde{\theta}^*. \end{aligned}$$

■

There is an explicit formula for the MLEs of  $\theta$  under the null hypothesis,

$$\hat{\theta} = (\hat{\theta}_{2(1)}, \hat{\theta}_{12(11)}, \dots, \hat{\theta}_{12(1I-1)})^T = (\text{logit}(\hat{\pi}_0), 0, \dots, 0)^T. \quad (18)$$

For the calculation of  $\tilde{\theta} = (\tilde{\theta}_{2(1)}, \tilde{\theta}_{12(11)}, \dots, \tilde{\theta}_{12(1I-1)})^T$  or  $\bar{\theta} = (\bar{\theta}_{2(1)}, \bar{\theta}_{12(11)}, \dots, \bar{\theta}_{12(1I-1)})^T$ , it is much easier to calculate first the corresponding MLE for the probability vector,  $\tilde{\pi} = (\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_I)^T$  or  $\bar{\pi} = (\bar{\pi}_1, \bar{\pi}_2, \dots, \bar{\pi}_I)^T$ , and plugging it to (11). There is an explicit formula for

$$\bar{\pi} = \left( \frac{N_{11}}{n_1}, \frac{N_{21}}{n_2}, \dots, \frac{N_{I1}}{n_I} \right)^T, \quad (19)$$

and for calculating  $\tilde{\pi}$  the following PAVA algorithm can be used.

**Algorithm 3 (Order restricted estimation of probabilities)** *The MLE of  $\pi = (\pi_1, \pi_2, \dots, \pi_I)^T$  under the restriction of  $\pi_1 \leq \pi_2 \leq \dots \leq \pi_I$ ,  $\tilde{\pi} = (\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_I)^T$ , is calculated in the following way:*

*STEP 1: Do  $\tilde{\pi} := \bar{\pi}$ , where  $\bar{\pi}$  is (19).*

*STEP 2: While not  $\tilde{\pi}_i \leq \tilde{\pi}_{i+1} \ \forall i = 1, \dots, I-1$  do*

*For  $i = 1, \dots, I-1$*

*If  $\tilde{\pi}_i \not\leq \tilde{\pi}_{i+1}$  do  $\tilde{\pi}_i := \frac{n_i}{n}\tilde{\pi}_i + \frac{n_{i+1}}{n}\tilde{\pi}_{i+1}$  and  $\tilde{\pi}_{i+1} := \tilde{\pi}_i$ .*

### 3 Phi-divergence test statistics

The classical order-restricted likelihood ratio test for testing (2) is given by

$$G^2 = 2 \sum_{i=1}^I \left( N_{i1} \log \frac{\tilde{\pi}_i}{\hat{\pi}_0} + (n_i - N_{i1}) \log \frac{1 - \tilde{\pi}_i}{1 - \hat{\pi}_0} \right)$$

(see for instance Mancuso et al (2001)). The Kullback-Leibler divergence measure between two  $2I$ -dimensional probability vectors  $\mathbf{p} = (p_{11}, p_{12}, \dots, p_{I1}, p_{I2})^T$  and  $\mathbf{q} = (q_{11}, q_{12}, \dots, q_{I1}, q_{I2})^T$ , is

given by

$$d_{Kull}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^I \left( p_{i1} \log \frac{p_{i1}}{q_{i1}} + p_{i2} \log \frac{p_{i2}}{q_{i2}} \right).$$

It is an easy exercise to verify that

$$G^2 = 2n(d_{Kull}(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}})) - d_{Kull}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}))), \quad (20)$$

where

$$\begin{aligned} \bar{\mathbf{p}} &= \mathbf{p}(\bar{\boldsymbol{\theta}}) = \left( \frac{n_1}{n} \bar{\pi}_1, \frac{n_1}{n} (1 - \bar{\pi}_1), \frac{n_2}{n} \bar{\pi}_2, \frac{n_2}{n} (1 - \bar{\pi}_2), \dots, \frac{n_I}{n} \bar{\pi}_I, \frac{n_I}{n} (1 - \bar{\pi}_I) \right)^T \\ &= \left( \frac{N_{11}}{n}, \frac{N_{12}}{n}, \frac{N_{21}}{n}, \frac{N_{22}}{n}, \dots, \frac{N_{I1}}{n}, \frac{N_{I2}}{n} \right)^T, \\ \tilde{\mathbf{p}} &= \mathbf{p}(\tilde{\boldsymbol{\theta}}) = \left( \frac{n_1}{n} \tilde{\pi}_1, \frac{n_1}{n} (1 - \tilde{\pi}_1), \frac{n_2}{n} \tilde{\pi}_2, \frac{n_2}{n} (1 - \tilde{\pi}_2), \dots, \frac{n_I}{n} \tilde{\pi}_I, \frac{n_I}{n} (1 - \tilde{\pi}_I) \right)^T, \\ \hat{\mathbf{p}} &= \mathbf{p}(\hat{\boldsymbol{\theta}}) = \left( \frac{n_1}{n} \hat{\pi}_0, \frac{n_1}{n} (1 - \hat{\pi}_0), \frac{n_2}{n} \hat{\pi}_0, \frac{n_2}{n} (1 - \hat{\pi}_0), \dots, \frac{n_I}{n} \hat{\pi}_0, \frac{n_I}{n} (1 - \hat{\pi}_0) \right)^T. \end{aligned}$$

The classical order-restricted chi-square test statistic for testing (2), known as Bartholomew's test-statistic, is given by

$$X^2 = \frac{1}{\hat{\pi}_0 (1 - \hat{\pi}_0)} \sum_{i=1}^I n_i (\tilde{\pi}_i - \hat{\pi}_0)^2, \quad (21)$$

which can be written as

$$X^2 = 2nd_{Pearson}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})), \quad (22)$$

where  $d_{Pearson}(\mathbf{p}, \mathbf{q})$  is the Pearson divergence measure defined by

$$d_{Pearson}(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \sum_{i=1}^I \left( \frac{(p_{i1} - q_{i1})^2}{q_{i1}} + \frac{(p_{i2} - q_{i2})^2}{q_{i2}} \right).$$

Details about this test-statistic can be found in Fleiss et al. (2003, Section 9.3).

More general than the Kullback-Leibler divergence and Pearson divergence measures are  $\phi$ -divergence measures, defined as

$$d_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^I \left( q_{i1} \phi \left( \frac{p_{i1}}{q_{i1}} \right) + q_{i2} \phi \left( \frac{p_{i2}}{q_{i2}} \right) \right),$$

where  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$  is a convex function such that  $\phi(1) = \phi'(1) = 0$ ,  $\phi''(1) > 0$ ,  $0\phi(\frac{0}{0}) = 0$ ,  $0\phi(\frac{p}{0}) = p \lim_{u \rightarrow \infty} \frac{\phi(u)}{u}$ , for  $p \neq 0$ . For more details about  $\phi$ -divergence measures see Pardo (2006).

Based on  $\phi$ -divergence measures we shall consider in this paper two families of order-



restricted  $\phi$ -divergence test statistics valid for testing (2) or (9). The first one generalizes the order-restricted likelihood ratio test given in (20) in the sense that we replace on it the Kullback-Leibler divergence measure by a phi-divergence measure and its expression is

$$\begin{aligned} T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) &= \frac{2n}{\phi''(1)} (d_\phi(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}})) - d_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}))) \\ &= \frac{2}{\phi''(1)} \left\{ \sum_{i=1}^I n_i \left( \hat{\pi}_0 \phi \left( \frac{N_{i1}}{n_i \hat{\pi}_0} \right) - \tilde{\pi}_i \phi \left( \frac{N_{i1}}{n_i \tilde{\pi}_i} \right) + (1 - \hat{\pi}_0) \phi \left( \frac{n_i - N_{i1}}{n_i (1 - \hat{\pi}_0)} \right) - (1 - \tilde{\pi}_i) \phi \left( \frac{n_i - N_{i1}}{n_i (1 - \tilde{\pi}_i)} \right) \right) \right\}. \end{aligned} \quad (23)$$

For  $\phi(x) = x \log x - x + 1$ , we get the likelihood ratio test.

The second one generalizes the order-restricted Pearson test statistic given in (22) in the sense that we replace on it the Pearson divergence measure by a phi-divergence measure and its expression is

$$\begin{aligned} S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) &= \frac{2n}{\phi''(1)} d_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \\ &= \frac{2}{\phi''(1)} \left\{ \sum_{i=1}^I n_i \left( \hat{\pi}_0 \phi \left( \frac{\tilde{\pi}_i}{\hat{\pi}_0} \right) + (1 - \hat{\pi}_0) \phi \left( \frac{1 - \tilde{\pi}_i}{1 - \hat{\pi}_0} \right) \right) \right\}. \end{aligned} \quad (24)$$

For  $\phi(x) = \frac{1}{2} (x - 1)^2$ , we get the Pearson test-statistics.

The following theorem provides the link between the both test-statistics,  $T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$  and  $S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ , and Wald-type test-statistics.

**Theorem 4** *For testing (2) or (9), the asymptotic distribution of*

$$T \in \{T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})), S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})), W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}), H(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}), D(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})\}$$

*is common and is given by*

$$\lim_{n \rightarrow \infty} \Pr(T \leq x) = \sum_{i=0}^{I-1} w_i(I-1, \mathbf{V}) \Pr(\chi_i^2 \leq x),$$

where

$$\mathbf{V} = \mathbf{G}_{I-1} \text{diag}^{-1}(\boldsymbol{\nu}^*) \mathbf{G}_{I-1}^T + \frac{1}{\nu_I} \mathbf{e}_{I-1} \mathbf{e}_{I-1}^T, \quad (25)$$

and  $\{w_i(I-1, \mathbf{V})\}_{i=0}^{I-1}$  is the set of weights such that  $\sum_{i=0}^{I-1} w_i(I-1, \mathbf{V}) = 1$  and its values are given in Theorem 5.

**Proof.** Let  $\hat{\boldsymbol{\theta}}$  be the  $I$ -dimensional vector given in (18). The second order Taylor expansion of

function  $d_\phi(\boldsymbol{\theta}) = d_\phi(\mathbf{p}(\boldsymbol{\theta}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$  about  $\hat{\boldsymbol{\theta}}$  is

$$d_\phi(\boldsymbol{\theta}) = d_\phi(\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \frac{\partial}{\partial \boldsymbol{\theta}} d_\phi(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} + \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} d_\phi(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + o\left(\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2\right), \quad (26)$$

where

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} d_\phi(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \mathbf{0}_{I-1}, \\ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} d_\phi(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \phi''(1) \mathcal{I}_F^{(n_1, \dots, n_I)}(\hat{\boldsymbol{\theta}}), \end{aligned}$$

and

$$\mathcal{I}_F^{(n_1, \dots, n_I)}(\boldsymbol{\theta}) = \mathbf{X}^T \text{diag}\left\{\frac{n_i}{n} \pi_i (1 - \pi_i)\right\}_{i=1}^I \mathbf{X}.$$

Let  $\bar{\boldsymbol{\theta}} = \mathbf{X}^{-1} \text{logit}(\bar{\boldsymbol{\pi}})$ , where  $\bar{\boldsymbol{\pi}}$  is (19). In particular, for  $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$  we have

$$d_\phi(\mathbf{p}(\bar{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = \frac{\phi''(1)}{2} (\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^T \mathcal{I}_F^{(n_1, \dots, n_I)}(\hat{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + o\left(\|\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\|^2\right)$$

and for  $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$

$$d_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = \frac{\phi''(1)}{2} (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^T \mathcal{I}_F^{(n_1, \dots, n_I)}(\hat{\boldsymbol{\theta}}) (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + o\left(\|\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\|^2\right),$$

and then taking into account that  $\lim_{n \rightarrow \infty} \mathcal{I}_F^{(n_1, \dots, n_I)}(\boldsymbol{\theta}) = \mathcal{I}_F(\boldsymbol{\theta})$ ,

$$\begin{aligned} T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) &= \frac{2n}{\phi''(1)} (d_\phi(\bar{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}})) - d_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}))) \\ &= n(\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^T \mathcal{I}_F(\hat{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) - n(\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^T \mathcal{I}_F(\tilde{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + o_P(1) \\ &= D(\bar{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) + o_P(1). \end{aligned}$$

In a similar way it is obtained

$$d_\phi(\mathbf{p}(\bar{\boldsymbol{\theta}}), \mathbf{p}(\tilde{\boldsymbol{\theta}})) = \frac{\phi''(1)}{2} (\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^T \mathcal{I}_F^{(n_1, \dots, n_I)}(\tilde{\boldsymbol{\theta}}) (\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + o\left(\|\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\|^2\right),$$

and then taking into account that  $\lim_{n \rightarrow \infty} \mathcal{I}_F^{(n_1, \dots, n_I)}(\boldsymbol{\theta}) = \mathcal{I}_F(\boldsymbol{\theta})$ ,

$$\begin{aligned} S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) &= \frac{2n}{\phi''(1)} d_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) \\ &= n(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^T \mathcal{I}_F(\hat{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + o_P(1) \\ &= H(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) + o_P(1) \end{aligned}$$

According to Proposition 4.4.1 in Silvapulle and Sen (2005)

$$G^2 = D(\bar{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) + o_P(1) = H(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) + o_P(1) = W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}) + o_P(1),$$

which means that the asymptotic distribution of

$$T \in \{T_\phi(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})), S_\phi(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})), W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}), H(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}), D(\bar{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})\}$$

is common. Such a distribution can be established from the likelihood ratio test-statistic used for the problem formulated in (6.13) of Silvapulle and Sen (2005)

$$\lim_{n \rightarrow \infty} \Pr(G^2 \leq x) = \sum_{i=0}^{I-1} w_i(I-1, \mathbf{V}) \Pr(\chi_i^2 \leq x),$$

where

$$\mathbf{V} = \mathbf{B} \text{diag}^{-1}(\boldsymbol{\nu}) \mathbf{B}^T,$$

with

$$\begin{aligned} \boldsymbol{\nu} &= (\boldsymbol{\nu}^*, \nu_I)^T = (\nu_1, \dots, \nu_I)^T, \\ \mathbf{B} &= (\mathbf{G}_{I-1}, -\mathbf{e}_{I-1}). \end{aligned}$$

Using the partitioned structure of  $\mathbf{B}$  and  $\boldsymbol{\nu}$ , (25) is obtained. ■

The following result is based on the third way for computation of weights given in page 79 of Silvapulle and Sen (2005).

**Theorem 5** *The set of weights  $\{w_i(I-1, \mathbf{V})\}_{i=0}^{I-1}$  of the asymptotic distribution given in Theorem 4 is computed as follows*

$$w_i(I-1, \mathbf{V}) = \Pr \left( \arg \min_{\boldsymbol{\zeta} \in \mathbb{R}_+^{I-1}} (\mathbf{Z} - \boldsymbol{\zeta})^T \mathbf{V}^{-1} (\mathbf{Z} - \boldsymbol{\zeta}) \in \mathbb{R}_+^{I-1}(i) \right), \quad i = 0, \dots, I-1 \quad (27)$$

$\mathbf{Z} \sim \mathcal{N}_{I-1}(\mathbf{0}_{I-1}, \mathbf{V})$ ,  $\mathbb{R}_+^{I-1} = \{\boldsymbol{\zeta} \in \mathbb{R}_+^{I-1} : \boldsymbol{\zeta} \geq \mathbf{0}\}$  and  $\mathbb{R}_+^{I-1}(i) \subset \mathbb{R}_+^{I-1}$  such that  $i$  components

are strictly positive and  $I - 1 - i$  components are null. In particular, for  $I \in \{2, 3, 4\}$  (27) has the explicit expressions given in (3.24), (3.25) and (3.26) of Silvapulle and Sen (2005).

For computing (27) is useful to know the following explicit expression  $\mathbf{V}^{-1} = \mathbf{T}_{I-1}^T \boldsymbol{\Sigma}_{\boldsymbol{\nu}^*} \mathbf{T}_{I-1}$ , where  $\mathbf{T}_{I-1} = \mathbf{G}_{I-1}^{-1}$  is an upper triangular matrix of 1-s,  $\boldsymbol{\Sigma}_{\boldsymbol{\nu}^*} = \text{diag}(\boldsymbol{\nu}^*) - \boldsymbol{\nu}^*(\boldsymbol{\nu}^*)^T$ , and to simulate the probability according to the following algorithm. For simulation

$$\hat{\mathbf{V}} = \mathbf{G}_{I-1} \text{diag}^{-1}(\hat{\boldsymbol{\nu}}^*) \mathbf{G}_{I-1}^T + \frac{1}{\hat{\nu}_I} \mathbf{e}_{I-1} \mathbf{e}_{I-1}^T, \quad (28)$$

$$\hat{\mathbf{V}}^{-1} = \mathbf{T}_{I-1}^T \boldsymbol{\Sigma}_{\hat{\boldsymbol{\nu}}^*} \mathbf{T}_{I-1}, \quad (29)$$

are needed rather than  $\mathbf{V}$  and  $\mathbf{V}^{-1}$  respectively.

**Algorithm 6 (Estimation of weights)** *The estimators of the weights given in Theorem 5,  $\hat{w}_i = w_i(I - 1, \hat{\mathbf{V}})$ , are obtained by Monte Carlo in the following way:*

*STEP 1: For  $i = 0, \dots, I - 1$ , do  $N(i) := 0$ .*

*STEP 2: Repeat the following steps  $R$  (say  $R = 1,000,000$ ) times:*

*STEP 2.1: Generate an observation,  $\mathbf{z}$ , from  $\mathbf{Z} \sim \mathcal{N}_{I-1}(\mathbf{0}_{I-1}, \hat{\mathbf{V}})$ . E.g., the NAG Fortran library subroutines G05CBF, G05EAF, and G05EZF can be useful.*

*STEP 2.2: Compute  $\hat{\zeta}(\mathbf{z}) = \arg \min_{\boldsymbol{\zeta} \in \mathbb{R}_+^{I-1}} \frac{1}{2} \boldsymbol{\zeta}^T \hat{\mathbf{V}}^{-1} \boldsymbol{\zeta} - (\hat{\mathbf{V}}^{-1} \mathbf{z})^T \boldsymbol{\zeta}$ . E.g., the IMSL Fortran library subroutine DQPROG can be useful.*

*STEP 2.3: Count  $i^*$ , the number of strictly positive components contained in  $\hat{\zeta}(\mathbf{z})$ , and*

*do  $N(i^*) := N(i^*) + 1$ .*

*STEP 3: Do  $\hat{w}_i := \frac{N(i)}{R}$  for  $i = 0, \dots, I - 1$ .*

## 4 Example

In this section the data set of the introduction (Table 1) is analyzed. The sample, a realization of  $\mathbf{N}$ , is summarized in the following vector

$$\begin{aligned} \mathbf{n} &= (n_{11}, n_{12}, n_{21}, n_{22}, n_{31}, n_{32}, n_{41}, n_{42})^T \\ &= (48, 17066, 38, 14464, 5, 788, 2, 163)^T. \end{aligned}$$

The estimated vectors of interest are

$$\begin{aligned}\bar{\pi} &= \left( \frac{48}{17114}, \frac{38}{14502}, \frac{5}{793}, \frac{2}{165} \right)^T = (0.0028, 0.0026, 0.0063, 0.0121)^T, \\ \tilde{\pi} &= \left( \frac{43}{15808}, \frac{43}{15808}, \frac{5}{793}, \frac{2}{165} \right)^T = (0.0027, 0.0027, 0.0063, 0.0121)^T, \\ \hat{\pi}_0 &= \frac{93}{32574} = 0.0029,\end{aligned}$$

and

$$\begin{aligned}\bar{\theta} &= \mathbf{X}^{-1} \text{logit}(\bar{\pi}) = (-4.4006, -1.473, -1.5412, -0.65946)^T, \\ \tilde{\theta} &= \mathbf{X}^{-1} \text{logit}(\tilde{\pi}) = (-4.4006, -1.5037, -1.5037, -0.65946)^T \\ \hat{\theta} &= (\text{logit}(\hat{\pi}_0), 0, 0, 0)^T = (-5.8558, 0, 0, 0)^T.\end{aligned}$$

For the asymptotic distribution the weighs can be calculated though

$$\hat{\mathbf{V}} = \begin{pmatrix} 83774.0156250 & -14428.7177734 & 0 \\ -14428.7177734 & 15217.7109375 & -788.9927979 \\ 0 & -788.9927979 & 1457.5666504 \end{pmatrix},$$

calculating the correlation coefficients

$$\hat{\rho}_{12} = -0.40411, \quad \hat{\rho}_{13} = 0, \quad \hat{\rho}_{23} = -0.16753,$$

the partial correlation coefficients

$$\hat{\rho}_{12 \bullet 3} = -0.4099, \quad \hat{\rho}_{13 \bullet 2} = -0.075072, \quad \hat{\rho}_{23 \bullet 1} = -0.18315,$$

and evaluating the following expressions

$$\begin{aligned}\hat{w}_3 &= \frac{1}{4\pi} (2\pi - \arccos(\hat{\rho}_{12}) - \arccos(\hat{\rho}_{13}) - \arccos(\hat{\rho}_{23})) = 0.07850, \\ \hat{w}_2 &= \frac{1}{4\pi} (3\pi - \arccos(\hat{\rho}_{12 \bullet 3}) - \arccos(\hat{\rho}_{13 \bullet 2}) - \arccos(\hat{\rho}_{23 \bullet 1})) = 0.32075, \\ \hat{w}_1 &= 0.5 - w_0(\hat{\theta}) = 0.5 - 0.07850 = 0.4215, \\ \hat{w}_0 &= 0.5 - w_1(\hat{\theta}) = 0.5 - 0.32075 = 0.17925.\end{aligned}$$

If we take  $\phi_\lambda(x) = \frac{1}{\lambda(1+\lambda)}(x^{\lambda+1} - x - \lambda(x-1))$ , where for each  $\lambda \in \mathbb{R} - \{-1, 0\}$ , the “power

divergence family” is obtained

$$d_\lambda(\mathbf{p}, \mathbf{q}) = \frac{1}{\lambda(\lambda + 1)} \left( \sum_{i=1}^I \sum_{j=1}^J \frac{p_{ij}^{\lambda+1}}{q_{ij}^\lambda} - 1 \right), \text{ for each } \lambda \in \mathbb{R} - \{-1, 0\}. \quad (30)$$

It is also possible to cover the real line for  $\lambda$ , by defining  $d_\lambda(\mathbf{p}, \mathbf{q}) = \lim_{t \rightarrow \lambda} d_t(\mathbf{p}, \mathbf{q})$ , for  $\lambda \in \{-1, 0\}$ . It is well known that  $d_0(\mathbf{p}, \mathbf{q}) = d_{Kull}(\mathbf{p}, \mathbf{q})$  and  $d_1(\mathbf{p}, \mathbf{q}) = d_{Pearson}(\mathbf{p}, \mathbf{q})$ . This is very interesting since this means that the power divergence based family of test-statistics contain as special cases  $G^2$  and  $X^2$ .

Finally, the expressions of the test-statistics are summarized in Table 2. It can be seen that the null hypothesis cannot be rejected for  $W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$ ,  $H(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$ ,  $D(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$ ,  $T_{-1.5}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $T_{-1}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $T_{-0.5}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $S_{-1.5}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $S_{-1}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $S_{-0.5}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$  and should be rejected for  $T_0(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $T_{\frac{2}{3}}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $T_1(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $S_0(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $S_{\frac{2}{3}}(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ ,  $S_1(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ . Even though the sample size seems to be large enough, this is a case of small values of  $\pi_1$ ,  $\pi_2$ ,  $\pi_3$ ,  $\pi_4$  which is known to have not reliable behavior in the values calculated for the  $p$ -values in order to make decisions. In the simulation study we shall study such a case and according to the results the rejection of the null hypothesis is supported since with  $\{T_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))\}_{\lambda \in \{0, \frac{2}{3}, 1\}}$  are obtained the most reliable test-statistics. As conclusion, an increase in maternal alcohol consumption is associated with an increase in the probability of malformation.

| test-statistic  | $\lambda = -1.5$ | $\lambda = -1$ | $\lambda = -0.5$ | $\lambda = 0$ | $\lambda = \frac{2}{3}$ | $\lambda = 1$ |
|---|------------------|----------------|------------------|---------------|-------------------------|---------------|
| $T_\lambda(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ | 3.3068           | 3.8173         | 4.4920           | 5.4057        | 7.2076                  | 8.4895        |
| $p\text{-value}(T_\lambda)$   | 0.1177           | 0.0911         | 0.0650           | 0.0413        | 0.0169                  | 0.0090        |
| $S_\lambda(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$                   | 3.2993           | 3.8124         | 4.4896           | 5.4057        | 7.2107                  | 8.4942        |
| $p\text{-value}(S_\lambda)$   | 0.1181           | 0.0913         | 0.0651           | 0.0413        | 0.0169                  | 0.0090        |

|   |        |   |        |  |        |
|---|--------|---|--------|--|--------|
| $W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$                 | 2.5979 | $H(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$                 | 2.6363 | $D(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$                 | 2.6462 |
| $p\text{-value}(W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}))$ | 0.1686 | $p\text{-value}(H(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}))$ | 0.1653 | $p\text{-value}(D(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}))$ | 0.1645 |

Table 2: Power divergence based test-statistics, Wald type statistics and their corresponding asymptotic  $p$ -values.

## 5 Simulation study

For testing (1), by considering  $I = 4$  binomial random variables, the following scenarios will be considered:

- scenario **A** (small/big proportions):  $n_1 = 40$ ,  $n_2 = 30$ ,  $n_3 = 20$ ,  $n_4 = 10$ .
- scenario **A-0**:  $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.05$ .

- scenario **A-1**:  $\pi_1 = 0.05, \pi_2 = \pi_3 = \pi_4 = 0.1$ .
- scenario **A-2**:  $\pi_1 = 0.05, \pi_2 = 0.1, \pi_3 = \pi_4 = 0.125$ .
- scenario **A-3**:  $\pi_1 = 0.05, \pi_2 = 0.1, \pi_3 = 0.125, \pi_4 = 0.135$ .
- scenario **B** (small/big proportions):  $n_1 = 60, n_2 = 45, n_3 = 30, n_4 = 15$ .
  - scenario **B-0**:  $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.05$ .
  - scenario **B-1**:  $\pi_1 = 0.05, \pi_2 = \pi_3 = \pi_4 = 0.1$ .
  - scenario **B-2**:  $\pi_1 = 0.05, \pi_2 = 0.1, \pi_3 = \pi_4 = 0.125$ .
  - scenario **B-3**:  $\pi_1 = 0.05, \pi_2 = 0.1, \pi_3 = 0.125, \pi_4 = 0.135$ .
- scenario **C** (small/big proportions):  $n_1 = 100, n_2 = 75, n_3 = 50, n_4 = 25$ .
  - scenario **C-0**:  $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.05$ .
  - scenario **C-1**:  $\pi_1 = 0.05, \pi_2 = \pi_3 = \pi_4 = 0.1$ .
  - scenario **C-2**:  $\pi_1 = 0.05, \pi_2 = 0.1, \pi_3 = \pi_4 = 0.125$ .
  - scenario **C-3**:  $\pi_1 = 0.05, \pi_2 = 0.1, \pi_3 = 0.125, \pi_4 = 0.135$ .
- scenario **D** (intermediate proportions):  $n_1 = 40, n_2 = 30, n_3 = 20, n_4 = 10$ .
  - scenario **D-0**:  $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.35$ .
  - scenario **D-1**:  $\pi_1 = 0.35, \pi_2 = \pi_3 = \pi_4 = 0.45$ .
  - scenario **D-2**:  $\pi_1 = 0.35, \pi_2 = 0.45, \pi_3 = \pi_4 = 0.475$ .
  - scenario **D-3**:  $\pi_1 = 0.35, \pi_2 = 0.45, \pi_3 = 0.475, \pi_4 = 0.485$ .
- scenario **E** (intermediate proportions):  $n_1 = 60, n_2 = 45, n_3 = 30, n_4 = 15$ .
  - scenario **E-0**:  $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.35$ .
  - scenario **E-1**:  $\pi_1 = 0.35, \pi_2 = \pi_3 = \pi_4 = 0.45$ .
  - scenario **E-2**:  $\pi_1 = 0.35, \pi_2 = 0.45, \pi_3 = \pi_4 = 0.475$ .
  - scenario **E-3**:  $\pi_1 = 0.35, \pi_2 = 0.45, \pi_3 = 0.475, \pi_4 = 0.485$ .
- scenario **F** (intermediate proportions):  $n_1 = 100, n_2 = 75, n_3 = 50, n_4 = 25$ .
  - scenario **F-0**:  $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.35$ .
  - scenario **F-1**:  $\pi_1 = 0.35, \pi_2 = \pi_3 = \pi_4 = 0.45$ .

- scenario **F-2**:  $\pi_1 = 0.35, \pi_2 = 0.45, \pi_3 = \pi_4 = 0.475$ .
- scenario **F-3**:  $\pi_1 = 0.35, \pi_2 = 0.45, \pi_3 = 0.475, \pi_4 = 0.485$ .

The simulation experiment is performed with  $R = 50000$  replications and in each of them, apart from the Wald type test-statistics  $T^{(h)} \in \{W^{(h)}(\tilde{\theta}, \hat{\theta}), H^{(h)}(\tilde{\theta}, \hat{\theta}), D^{(h)}(\tilde{\theta}, \tilde{\theta}, \hat{\theta})\}$ ,  $h = 1, \dots, R$ , all the power divergence test statistics,  $T^{(h)} \in \{T_{\lambda}^{(h)}(\bar{p}, p(\tilde{\theta}), p(\hat{\theta})), S_{\lambda}^{(h)}(p(\tilde{\theta}), p(\hat{\theta}))\}_{\lambda \in I}$ ,  $h = 1, \dots, R$ , associated with the interval  $I = (-1.5, 3)$  are considered. From the p-values it is possible to calculate the proportion of replications rejected according with the nominal size  $\alpha = 0.05$ , i.e.

$$\frac{\sum_{h=1}^R I\{p\text{-value}(T^{(h)}) \leq \alpha\}}{R}, \quad (31)$$

where  $I\{\bullet\}$  represents the indicator function. The scenarios ending in 0 represent that the null hypothesis is true, and are useful to obtain the simulated significance levels,  $\hat{\alpha}_T$ ,  $T \in \{W(\tilde{\theta}, \hat{\theta}), H(\tilde{\theta}, \hat{\theta}), D(\tilde{\theta}, \tilde{\theta}, \hat{\theta}), T_{\lambda}(\bar{p}, p(\tilde{\theta}), p(\hat{\theta})), S_{\lambda}(p(\tilde{\theta}), p(\hat{\theta}))\}_{\lambda \in I}$ , with different sample sizes and kinds of test-statistics. The scenarios ending in either 1, 2 or 3 represent that the null hypothesis is false and are useful to obtain the simulated powers,  $\hat{\beta}_T$ , with different alternatives, sample sizes and types of test-statistics. For calculating both,  $\hat{\alpha}_T$  and  $\hat{\beta}_T$ , (31) is applied, each one in the corresponding scenario.

In Figures 1 and 2,  $\hat{\alpha}_T$  and  $\hat{\beta}_T$  for all the aforementioned test-statistics are plotted in different scenarios. The curves represent either  $T_{\lambda}(\bar{p}, p(\tilde{\theta}), p(\hat{\theta}))$  or  $S_{\lambda}(p(\tilde{\theta}), p(\hat{\theta}))$  power divergence test-statistics, located respectively on left or right of the panel of plots. The asterisk, square and circle symbols, represent  $W(\tilde{\theta}, \hat{\theta})$ ,  $H(\tilde{\theta}, \hat{\theta})$  and  $D(\tilde{\theta}, \tilde{\theta}, \hat{\theta})$  respectively and all of them are repeated on the left as well as on the right in order to make easier their comparison with  $T_{\lambda}(\bar{p}, p(\tilde{\theta}), p(\hat{\theta}))$  or  $S_{\lambda}(p(\tilde{\theta}), p(\hat{\theta}))$  respectively. The black color lines and symbols, representing  $\hat{\alpha}_T$ , are useful to select the test-statistics closed to nominal level. According to the criterion given by Dale (1986), a reasonable exact significance level should verify

$$|\text{logit}(1 - \hat{\alpha}_T) - \text{logit}(1 - \alpha)| \leq \epsilon$$

for  $\epsilon \in \{0.35, 0.7\}$ , being “closed to nominal level” the exact significance levels verifying the inequality with  $\epsilon = 0.35$  and “fairly closed to nominal level” the ones with  $\epsilon = 0.7$ . In this study only the test statistics satisfying the condition with  $\epsilon = 0.35$  are considered, and the corresponding upper and lower bounds appear plotted with two horizontal lines, having in the middle the line associated with the nominal level,  $\alpha = 0.05$ . Among the test-statistics with simulated significance levels closed to the nominal level, the test-statistics with higher powers should be selected but since in general high powers correspond to high significance levels, this choice is not straightforward. For this reason, based on  $\hat{\beta}_{T_0} - \hat{\alpha}_{T_0}$  or  $\hat{\beta}_{S_1} - \hat{\alpha}_{S_1}$  as baseline, the



efficiencies relative to the likelihood ratio test  $(T_0(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = G^2)$

$$\rho_T = \frac{(\hat{\beta}_T - \hat{\alpha}_T) - (\hat{\beta}_{T_0} - \hat{\alpha}_{T_0})}{\hat{\beta}_{T_0} - \hat{\alpha}_{T_0}} \quad (32)$$

or the Bartholomew's test  $(S_1(\mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}})) = X^2)$

$$\rho_T^* = \frac{(\hat{\beta}_T - \hat{\alpha}_T) - (\hat{\beta}_{S_1} - \hat{\alpha}_{S_1})}{\hat{\beta}_{S_1} - \hat{\alpha}_{S_1}} \quad (33)$$

are considered. In Table 3, the efficiency of  $S_1 = X^2$  is compared with respect to  $T_0 = G^2$ , and then if  $\rho_{S_1} < 0$ , since  $T_0 = G^2$  is better than  $S_1 = X^2$ , the plot of the efficiencies in Figures 3 and 4 will be only focussed on (32). Similarly, if  $\rho_{S_1} > 0$ , since  $T_0 = G^2$  is worse than  $S_1 = X^2$ , the plot of the efficiencies in Figures 3 and 4 will be only focussed on (33).

|   | sc A             | sc B             | sc C             | sc D             | sc E             | sc F             |
|---|------------------|------------------|------------------|------------------|------------------|------------------|
| 1 | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ |
| 2 | $\rho_{S_1} < 0$ | $\rho_{S_1} > 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ |
| 3 | $\rho_{S_1} < 0$ | $\rho_{S_1} > 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ | $\rho_{S_1} < 0$ |

Table 3: Efficiency of the Bartholomew's test with respect to the likelihood ratio test.

In view of the plots, it is possible to propose test-statistics with better performance in comparison with  $G^2$  and  $X^2$ . From Figures 1 and 3, the so-called Cressie-Read test-statistic,  $T_{\frac{2}{3}}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ , can be recommended for small/big proportions either for small or moderate sample sizes. On the other hand,  $W(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$ ,  $H(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$  and the test-statistic based on the Hellinger distance,  $T_{-0.5}(\bar{\mathbf{p}}, \mathbf{p}(\tilde{\boldsymbol{\theta}}), \mathbf{p}(\hat{\boldsymbol{\theta}}))$ , can be recommended for intermediate proportions and moderate sample sizes, however for small sample sizes the likelihood ratio test-statistic still remains being the best one.

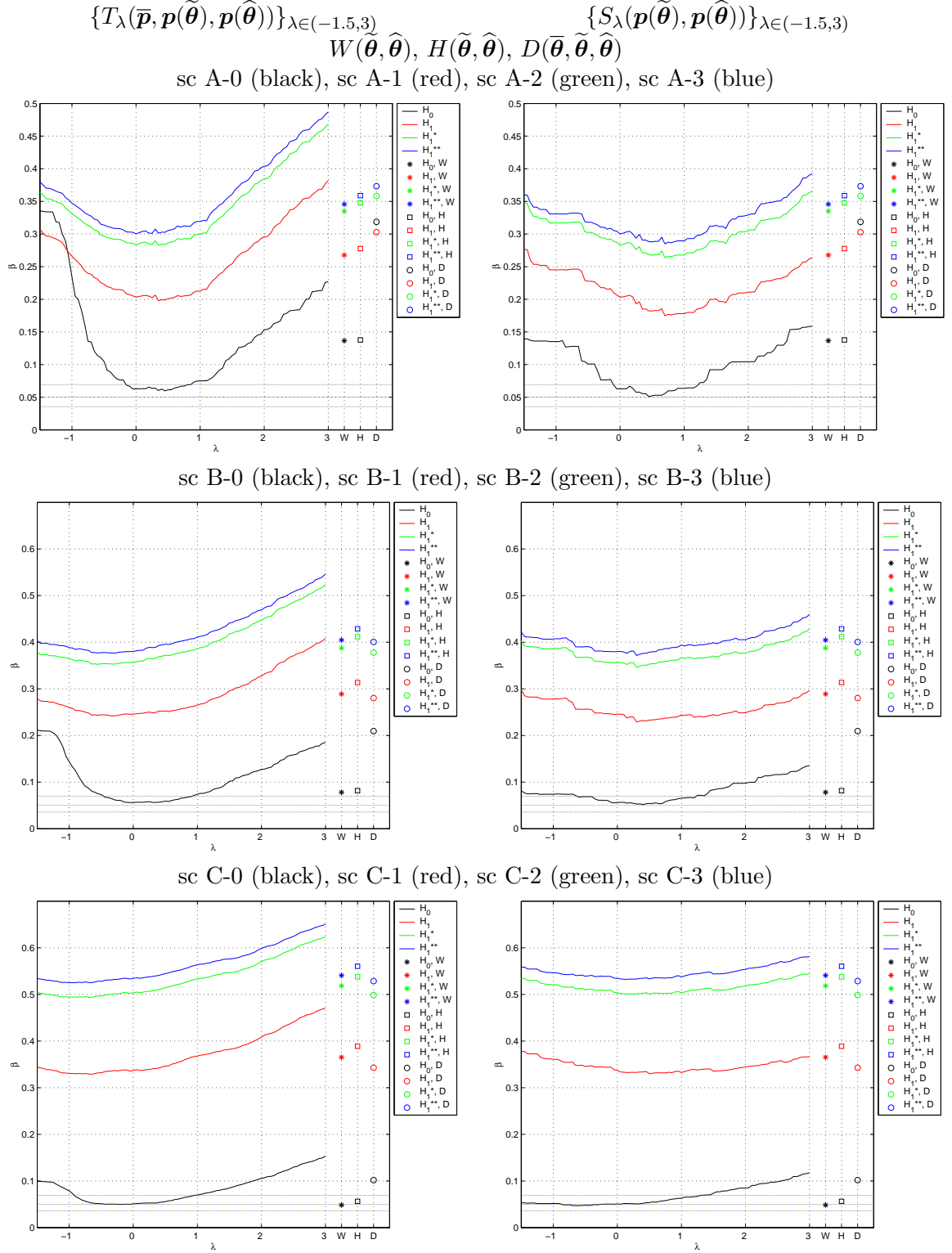


Figure 1: Simulated sizes (black) and powers (red, green, blue) for scenarios A,B,C (small/big proportions).

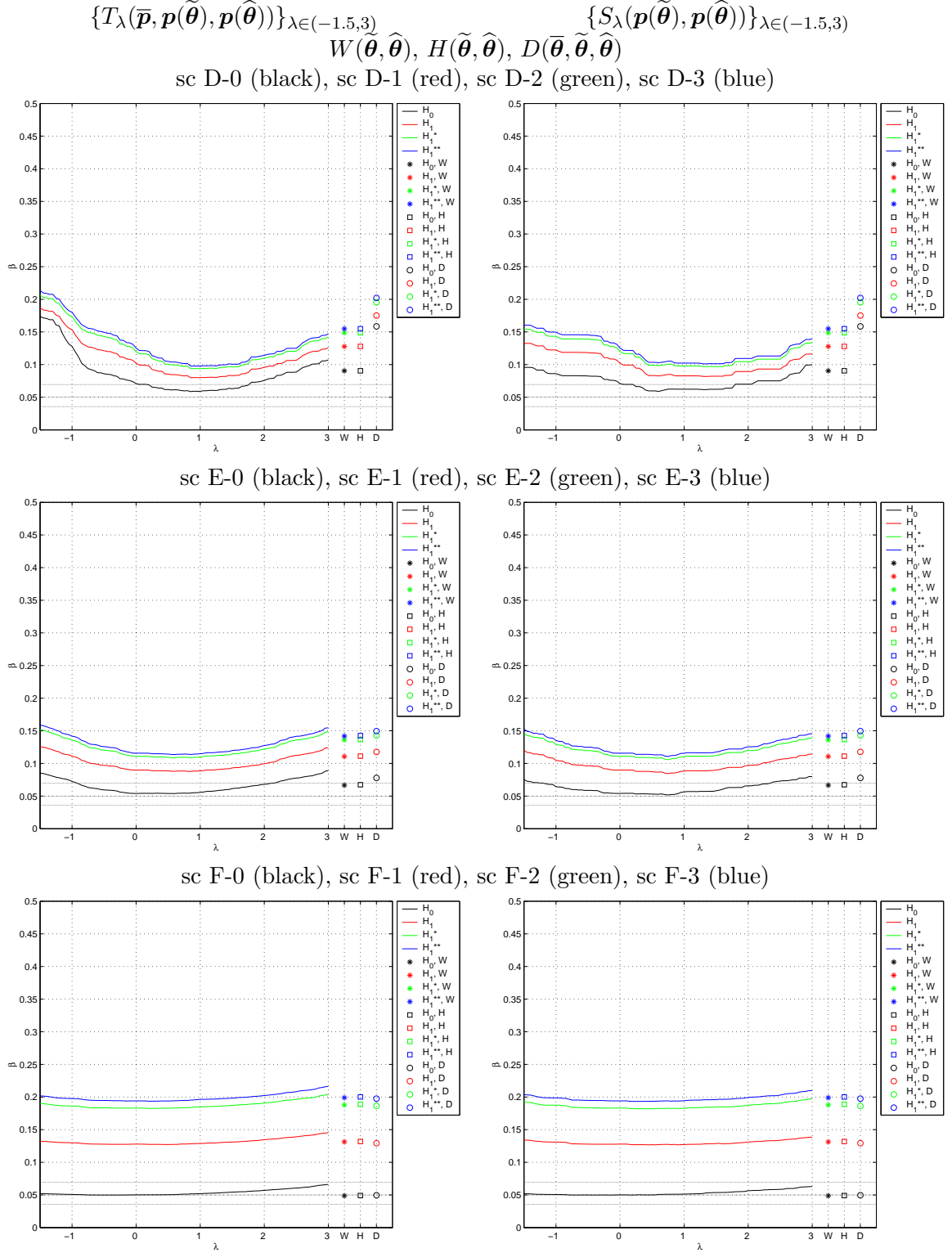


Figure 2: Simulated sizes (black) and powers (red, green, blue) for scenarios D,E,F (intermediate proportions).

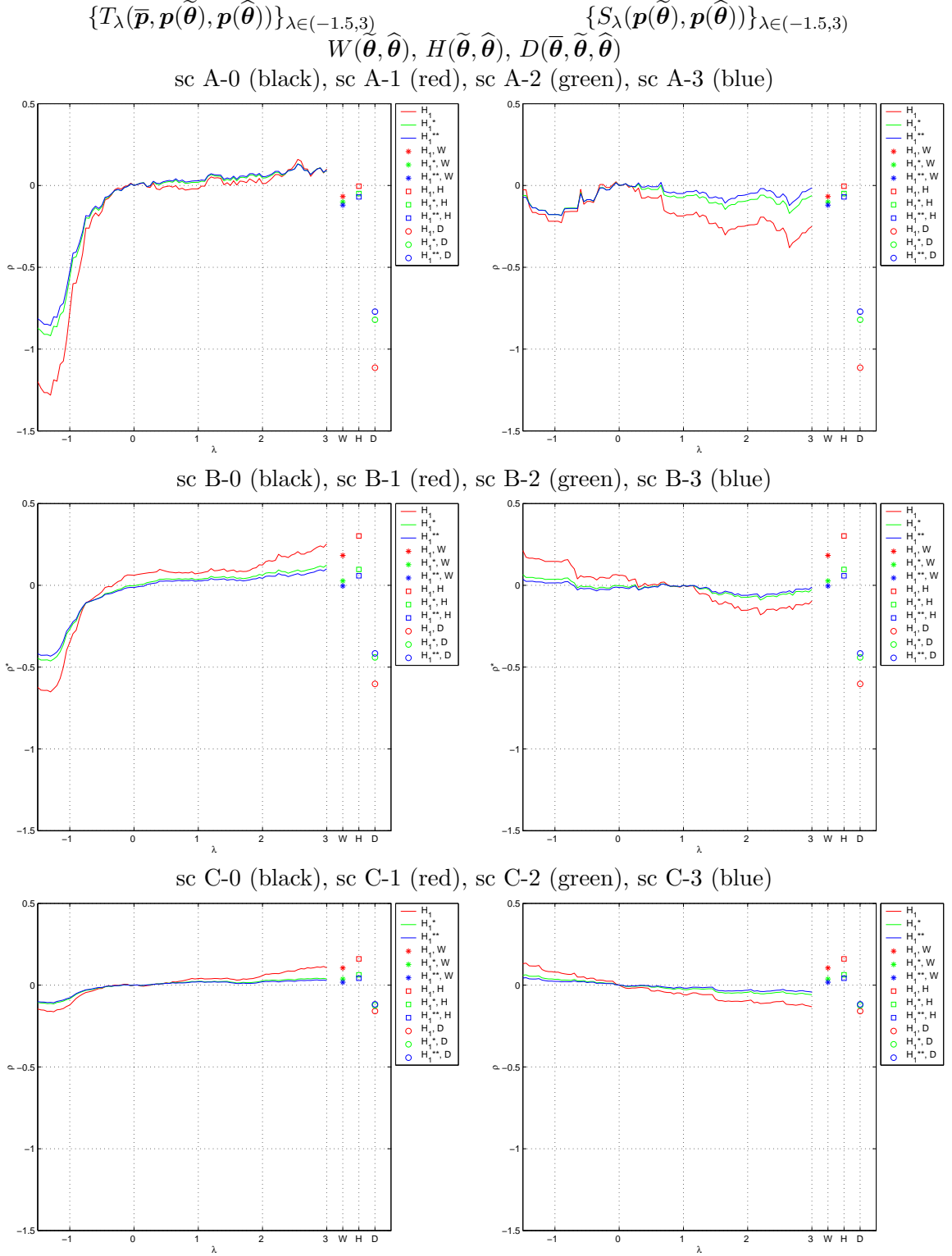


Figure 3: Efficiencies for scenarios A,B,C (small/big proportions).

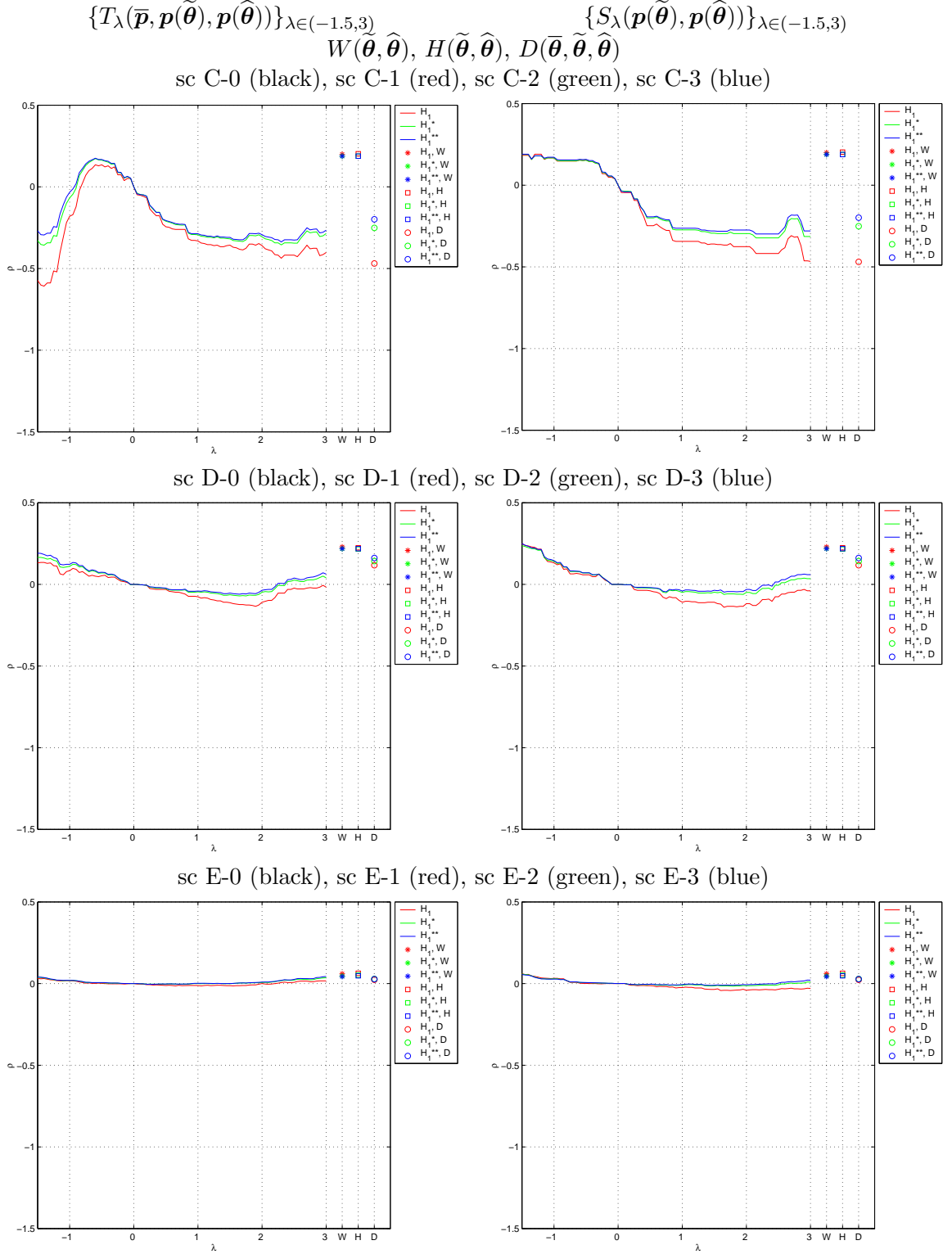


Figure 4: Efficiencies for scenarios C,D,E (intermediate proportions).

## References

- [1] BARLOW, R. E., BARTHOLOMEW, D. J. AND BRUNK, H.D. (1972). *Statistical inference under order restrictions*. Wiley.
- [2] DALE, J.R. (1986). Asymptotic normality of goodness-of-fit statistics for sparse product multinomials. *Journal of the Royal Statistical Society*, **B**, **48**, 48–59.
- [3] DARDANONI, V. AND FORCINA, A. (1998). A Unified Approach to Likelihood Inference on Stochastic Orderings in a Nonparametric Context. *Journal of the American Statistical Association*, **93**, 1112–1122.
- [4] FLEISS, J.L., LEVIN B. AND PAIK, M.C. (2003). *Statistical Methods for Rates and Proportions*. Wiley Interscience.
- [5] GRAUBARD, B. I. AND KORN, E. L. (1987). Choice of Column Scores for Testing Independence in Ordered  $2 \times I$  Contingency Tables. *Biometrics*, **43**, 471-476.
- [6] MARTIN, N. AND PARDO, L. (2008). New families of estimators and test statistics in log-linear models. *Journal of Multivariate Analysis*, **99**(8), 1590-1609.
- [7] MARTIN, N., MATA, R. AND PARDO, L. (2014). Phi-divergence statistics for the likelihood ratio order: an approach based on log-linear models. <http://arxiv.org/pdf/1402.5384v1.pdf>.
- [8] MANCUSO, J. Y.; AHAN, H. AND CHEN, J. J. (2001). Order-restricted dose-related trend tests. *Statistics in Medicine*, **20**, 2305-2318.
- [9] PARDO, L. (2006). *Statistical Inference Based on Divergence Measures*. Statistics: series of Textbooks and Monographs. Chapman & Hall / CRC.
- [10] SILVAPULLE, M. J. AND SEN., P. K. (2005). *Constrained statistical inference. Inequality, order, and shape restrictions*. Wiley Series in Probability and Statistics. Wiley-Interscience (John Wiley & Sons).